

**Parametric Evaluation of Different ANN Architectures  
Forecasting Wind Power Across Different Time Horizons**

Sewdien, V. N.; Preece, R.; Rueda Torres, J. L.; Van Der Meijden, M. A.M.M.

**DOI**

[10.1109/TDC-LA.2018.8511757](https://doi.org/10.1109/TDC-LA.2018.8511757)

**Publication date**

2018

**Document Version**

Accepted author manuscript

**Published in**

Proceedings of the 2018 IEEE PES Transmission and Distribution Conference and Exhibition - Latin America, T and D-LA 2018

**Citation (APA)**

Sewdien, V. N., Preece, R., Rueda Torres, J. L., & Van Der Meijden, M. A. M. M. (2018). Parametric Evaluation of Different ANN Architectures: Forecasting Wind Power Across Different Time Horizons. In *Proceedings of the 2018 IEEE PES Transmission and Distribution Conference and Exhibition - Latin America, T and D-LA 2018* Article 8511757 IEEE. <https://doi.org/10.1109/TDC-LA.2018.8511757>

**Important note**

To cite this publication, please use the final published version (if applicable).  
Please check the document version above.

**Copyright**

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**

Please contact us and provide details if you believe this document breaches copyrights.  
We will remove access to the work immediately and investigate your claim.

# Evaluation of Artificial Neural Network Parameters: Optimization for Increased Wind Power Forecast Accuracy

V.N. Sewdien  
TenneT TSO B.V.  
Arnhem, The Netherlands  
Vinay.Sewdien@tennet.eu

R. Preece  
School of Electrical &  
Electronic Engineering  
University of Manchester  
Manchester, UK

J.L. Rueda Torres  
Intelligent Electrical  
Power Grids  
Delft University of  
Technology  
Delft, The Netherlands

M.A.M.M. van der  
Meijden  
Intelligent Electrical  
Power Grids  
Delft University of  
Technology  
Delft, The Netherlands

**Abstract**— The participation of volatile wind energy in the generation mix of the power system is increasing. . It becomes crucial for system operators to accurately predict the wind power generation across different short term horizons. This research aims to investigate how different parameters of artificial neural networks, such as the input data, batch size, number of hidden layers, number of neurons per hidden layer, and the amount of training data, influence the short term forecast accuracy. From the investigated cases it is concluded that a neural network with two hidden layers always has the best performance. Furthermore, with increasing forecast horizons, better performance is achieved when neural networks contain increased neurons in the hidden layers and have enlarged training data sets.

**Index Terms**— Wind Energy, Forecasting, Artificial Neural Networks, Sensitivity.

## I. INTRODUCTION

With increasing penetration of wind generation it becomes essential for system operators to accurately predict the wind power, in order to ensure reliable and affordable supply of electricity. This forecasting is done across different time horizons. For different forecast horizons (FH) there exist different models: statistical models for up to 6 hours ahead and numerical weather prediction (NWP) based physical models for more than 6 hours ahead. For statistical models, artificial neural networks (ANN) are one of the most accurate techniques. This research focuses on ANN based statistical models for forecast horizons of 5, 15, 30, and 60 minutes ahead. The 5 minutes forecast horizon (FH 5) is useful for ramp forecasting, which is crucial for power systems with high penetration of wind generation [1][2][3]. An emergency event in Texas related to the wind power ramping is described in [4]. FH 15 and FH 60 are useful for intraday markets where quarter-hourly and hourly products are traded.

The aim of this research is to investigate how the forecast accuracy across the different horizons is influenced by changes in the following parameters of the ANN:

- Number of inputs, i.e. the historic data set (HDS): 5, 10, 20.
- Number of hidden layers (HL): 1, 2, 3.
- Number of neurons per hidden layer ( $N_{HL}$ ): 100% (i.e. equal to the number of neurons in the input layer) and 50% (i.e. equal to the average of the neurons in the input and output layer).
- Size of the training data set (TDS): 50% and 80%.
- Batch size (BS), i.e. amount of observations after which the weighting factors are updated: 5, 10, 20.

Whereas a majority of the publications investigated the influence of the HDS on the forecast accuracy, only some did analyze the impact of the HDS combined with one or more aspects of the ANN's structure. In one study the influence of the HDS size for a single 1 hour forecast was investigated. The forecasting algorithm contained 1 hidden layer with 3 neurons, with TDS 57%. It was found that the optimum size of the HDS is dependent on the learning rate of the algorithm [5]. In another study the influence of HDS on the forecast accuracy in terms of root mean square error for FH 30 is presented. The implemented forecasting algorithm contained 1 hidden layer, whereas the HDS was varied from 3 to 8. It was concluded that the highest forecast accuracy is achieved for the ANN with HDS 8 [6]. In [7] the influence of HL and HDS on the forecast accuracy was investigated. It was found that a simple ANN with HDS 2 and no hidden layers performed the best in terms of forecast accuracy.

The current research investigates the combined influence of the HDS and certain aspects of the ANN's structure (not only limited to the number of HL) on the forecast accuracy for four different forecast horizons.

## II. RESEARCH METHOD

As stated earlier, the aim of this research is to examine to what extent certain parameters and settings of the ANN influence the accuracy of wind power forecasts (the focus is not on minimizing the forecast error). This influence is investigated for four different forecast horizons.

### A. Artificial Neural Network

Figure 1 illustrates the general architecture of an ANN.

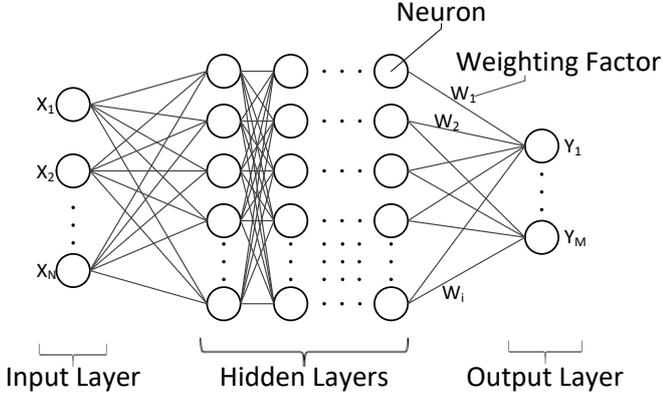


Fig. 1 Artificial Neural Network: General Architecture

It consists of an input layer, one or more hidden layers, an output layer, and several synapses with their associated weighting factors. Each layer contains a number of neurons. For the input layer this can be either previous observed values of the wind power generation or numerical weather prediction data such as wind speed, pressure, and temperature. A single neuron in the input layer is assigned to each input variable. The number of neurons in the hidden layers can be chosen arbitrarily. An activation function is used to determine the weighting factors of the neurons in the last hidden layer. The dimension of the output layer is determined by the number of outputs being forecasted. Based on the objective function of the ANN's optimizer, the weighting factors are updated using the backward propagation technique [8].

The ANN developed for this paper is modelled in Python [9]. The parameters that were kept constant during the analysis are given in Table 1.

TABLE I. FIXED PARAMETERS ANN

Parameter	Value
Epochs	100
Activation Function	Rectifier Linear Unit
Output Layer Neurons	1
Optimizer	Adam [10]

The FH for which analyses were carried out are 5, 15, 30, and 60 minutes ahead. For these forecast horizons better accuracies are achieved when instead of NWP data, past observed wind power generation values are used as input [11]–[13]. In total 27 cases combining several ANN parameters were investigated

(see Appendix A). The characteristics of the Base Case are given in Table 2.

TABLE II. BASE CASE VARIABLES

Variable	Value
HDS	10
HL	2
N <sub>HL</sub>	50%
TDS	80%
BS	10

### B. Data

The data used for this research was retrieved from the WIND Prospector Toolkit of USA's National Renewable Energy Laboratory and belongs to a small wind park of 16 MW (Site ID 8501) [14]–[17]. Observed NWP data (wind direction, wind speed, air temperature, surface air pressure, and air density) and wind active power generation data with a resolution of 5 minutes are available for the time span 2007–2012. The statistical parametric t test was performed successfully on the data sets to determine if all the data belonged to the same population.

### C. Forecast Error: Mean Absolute Error

In order to assess the influence of the variables on the forecast accuracy, literature provides measures such as root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). For the reasons given in [11], [18], [19] the normalized MAE (nMAE) will be used as measure for the forecast accuracy. The nMAE is calculated as:

$$nMAE = \frac{1}{P_{MAX}} \frac{1}{n} \sum_{i=1}^n |y_i - y_{ip}|, \quad (1)$$

where  $P_{MAX}$  is the maximum active power generation of the wind farm,  $n$  is the number of observations,  $y_i$  is the observed wind generation for instance  $i$ , and  $y_{ip}$  is the forecasted wind generation for instance  $i$ .

For each of the 108 cases (27 cases per forecasting horizon), the ANN is trained using the data from 2007. After the training, the ANN is evaluated by calculating the nMAE for each year of data (2008 – 2012). The final nMAE reported per case in this paper is the average nMAE over the 5 years of that case. An example for case 15 for FH 5 is given in Table 3.

TABLE III. nMAE FOR CASE 15 (FH=5)

Year	nMAE
2008	2.778%
2009	2.398%
2010	2.622%
2011	2.442%
2012	2.469%

Year	nMAE
Final nMAE	2.542%

### III. RESULTS & DISCUSSION

In Figure 2 the forecast error distribution is given for FH is 5, 15, 30, and 60 minutes. Accuracies on the outer circle of 0.3 need to be disregarded: the ANN in these cases did not give any output. From the figure it can be observed that the developed forecast algorithms have a low bias across all the investigated horizons. On the other hand, with increasing forecast horizon, the general trend observed is one of an increasing variance. Low bias-low variance algorithms are preferred. There is, however, always a tradeoff between the bias and the variance.

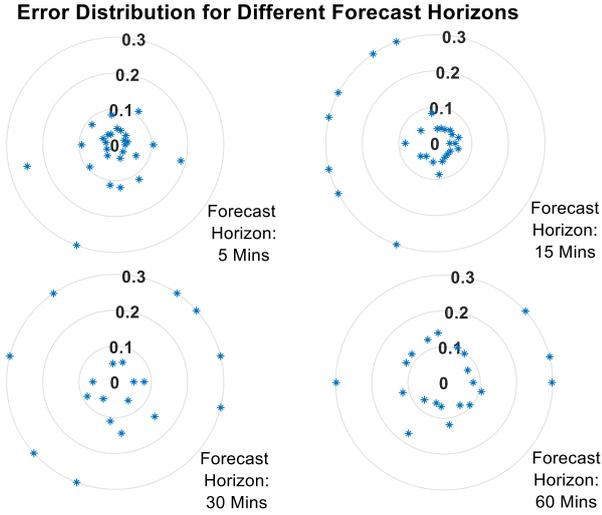


Fig. 2 Error Distribution for Different Forecast Horizons

In Figure 3 the nMAE of the forecasting algorithms is given for all the cases across the four forecasting horizons.

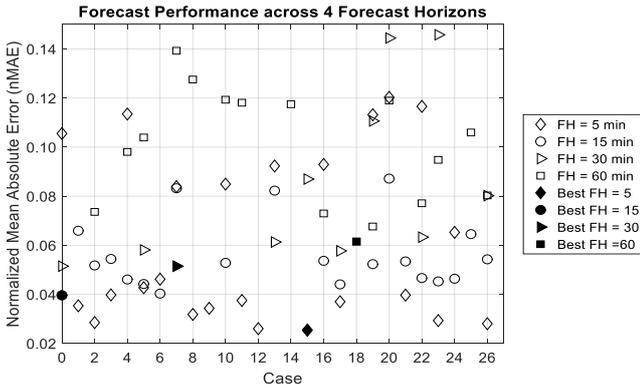


Fig. 3 Forecast Performance across Four Different Forecast Horizons

#### A. Forecast Horizon: 5 Minutes

For this FH, the general trend observed is increasing forecast accuracies (i.e. a decreasing nMAE) with increasing dimension of the HDS. This observation is independent on NHL and BS. In 8 of the 9 cases, an ANN with  $N_{HL}$  50%

outperforms an ANN with  $N_{HL}$  100%. Only with HDS 5, the ANN with HL 1 performs better. In all the other cases, the ANNs with HL 2 have a higher accuracy. The general trend is that with increasing dimension of HDS, cases with TDS 80% result in a slightly better performance. As the number of inputs is lower in the case with HDS 5, the ANN can be trained relatively better with less data. Therefore the case with TDS 50% and HDS 5 achieves a higher performance. When combining the variables, the best performance is achieved for an ANN with HL 2,  $N_{HL}$  50%, HDS 5, TDS 50% and BS 20. The average nMAE over the 5 years is 2.54%.

#### B. Forecast Horizon: 15 Minutes

The observation is that with HDS 5, the highest accuracy is achieved for BS 20. With HDS 20 and BS 5 the highest accuracy is achieved for the case with  $N_{HL}$  50%. With HDS 20 and BS 10 the highest accuracy is achieved for  $N_{HL}$  100%. In 6 out of the 9 cases, an ANN with  $N_{HL}$  50% outperforms an ANN with  $N_{HL}$  100%. In the other 3 cases,  $N_{HL}$  100% results in a slightly lower nMAE. With HDS 5, the ANN with HL 2 performs the best. With HDS 10, the ANN with HL 3 layers performs best. With HDS 20, no reliable result is achieved. For HDS 5, the best performance is achieved for TDS 80%. For HDS 10 and HDS 20, lowest nMAE achieved for TDS 50%. When combining the variables, the best performance is achieved for an ANN with HL 2,  $N_{HL}$  50%, HDS 10, TDS 80% and BS 10. The average nMAE over the 5 years is 3.96%.

#### C. Forecast Horizon: 30 Minutes

In 4 out of 6 cases, HDS 10 resulted in a better performance. In all the cases  $N_{HL}$  50% outperforms ANN with  $N_{HL}$  100%. The best performance is achieved for ANN with HL 2. For HDS 5 and HDS 20 best performance is achieved with TDS 50%. For HDS 10, best performance is achieved with TDS 80%. After combining various values of the parameters, the best performance is achieved for an ANN with HL 2,  $N_{HL}$  50%, HDS 10, TDS 80% and BS 10. The average nMAE over the 5 years is 5.15%.

#### D. Forecast Horizon: 60 Minutes

In terms of the batch size, the best performance is achieved for BS 5. When HDS 5, the lowest nMAE is achieved with  $N_{HL}$  100%. For HDS 10 and HDS 20,  $N_{HL}$  50% results in a higher accuracy. When varying the number of hidden layers, it is observed that an ANN with HL 3 outperforms ANNs with HL 2 or HL 1. Also, with increasing HDS, a higher TDS leads to increased accuracy. The overall best performance, when combining the various parameters, is achieved for an ANN with HL 2,  $N_{HL}$  100%, HDS 5, TDS 80% and BS 5. The average nMAE over the 5 years is 6.15%.

The variables for the best performing ANN for each FH are given in Table 4.

TABLE IV. VARIABLES OF BEST PERFORMING ANN

FH	Case	HL	$N_{HL}$	BS	TDS	HDS	nMAE
5	15	2	50%	20	50%	5	2.54%

FH	Case	HL	N <sub>HL</sub>	BS	TDS	HDS	nMAE
15	0	2	50%	10	80%	10	3.96%
30	7	2	50%	20	80%	10	5.15%
60	18	2	80%	5	80%	5	6.15%

From the 27 investigated cases per forecasting horizon, it can be concluded that the best performance is achieved when the neural network contains two hidden layers, independent of the forecast horizon. Furthermore, with increasing forecast horizons, better performance is achieved when the neural networks contain increased neurons in the hidden layers and have enlarged training data sets.

#### IV. CONCLUSIONS

The aim of this research was to investigate to what extent certain parameters and settings of an artificial neural network influence the accuracy of wind power forecasts across four forecast horizons: 5, 15, 30, and 60 minutes ahead. The results presented in this paper are based on 27 specific cases for each of the four forecast horizons. From these investigated cases it is observed that with increasing forecast horizons the variance of the forecast accuracy is increasing, whereas the bias remains low.

Furthermore, it can be concluded that the best performance is achieved when the neural network contains two hidden layers, independent of the forecast horizon. Furthermore, with increasing forecast horizons, better performance is achieved when neural networks contain increased neurons in the hidden layers and have enlarged training data sets. The influence of the batch size and the historic data size on the forecast accuracy are dependent on the structure of the artificial neural network.

As the influence of several parameters on the forecast performance is now known, an optimization of the most influencing parameters can be carried out with the aim to minimize the forecast error.

#### ACKNOWLEDGMENT

This research was carried out as part of the MIGRATE project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 691800. This paper reflects only the authors' views and the European Commission is not responsible for any use that may be made of the information it contains

#### REFERENCES

- [1] J. Zhang, M. Cui, B. M. Hodge, A. Florita, and J. Freedman, "Ramp forecasting performance from improved short-term wind power forecasting over multiple spatial and temporal scales," *Energy*, vol. 122, pp. 528–541, 2017.
- [2] A. Botterud, J. Wang, V. Miranda, and R. J. Bessa, "Wind Power Forecasting in U.S. Electricity Markets," *Electr. J.*, vol. 23, no. 3, pp. 71–82, 2010.
- [3] N. Cutler, M. Kay, H. Outhred, and I. MacGill, "High-Risk Scenarios for Wind Power Forecasting in Australia," 2007 *Eur. Wind Energy Conf. Exhib.*, 2007.
- [4] E. Ela and B. Kirby, "ERCOT Event on February 26 , 2008 : Lessons Learned," 2008.
- [5] G. Li and J. Shi, "On comparing three artificial neural networks for wind speed forecasting," *Appl. Energy*, vol. 87, no. 7, pp. 2313–2320, 2010.
- [6] M. Monfared, H. Rastegar, and H. M. Kojabadi, "A new strategy for wind speed forecasting using artificial intelligent methods," *Renew. Energy*, vol. 34, no. 3, pp. 845–848, 2009.
- [7] J. P. S. Catalão, H. M. I. Pousinho, and V. M. F. Mendes, "Short-term wind power forecasting in Portugal by neural networks and wavelet transform," *Renew. Energy*, vol. 36, no. 4, pp. 1245–1251, 2011.
- [8] M. A. Nielsen, *Neural Networks and Deep Learning*. Determination Press, 2015.
- [9] "Python Language Reference." Python Software Foundation.
- [10] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in *International Conference on Learning Representations*, 2014, pp. 1–15.
- [11] G. Giebel, R. Brownsword, G. Kariniotakis, M. Denhard, and C. Draxi, "Report: The state of the art in short-term prediction of wind power," p. 110, 2011.
- [12] G. Giebel and G. Kariniotakis, "Best Practice in Short-Term Forecasting. A Users Guide," in *European Wind Energy Conference and Exhibition 2007*, 2007.
- [13] G. Giebel and G. Kariniotakis, *Wind power forecasting—a review of the state of the art*, no. 2007. Elsevier Ltd, 2017.
- [14] C. Draxl and A. C. Nrel, "A Guide to Using the WIND Toolkit Validation Code," 2014.
- [15] C. Draxl, A. Clifton, B. M. Hodge, and J. McCaa, "The Wind Integration National Dataset (WIND) Toolkit," *Appl. Energy*, vol. 151, pp. 355–366, 2015.
- [16] C. Draxl, B. Hodge, and A. Clifton, "Overview and Meteorological Validation of the Wind Integration National Dataset Toolkit," 2015.
- [17] J. King, A. Clifton, and B.-M. Hodge, "Validation of Power Output for the WIND Toolkit," 2014.
- [18] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *Int. J. Forecast.*, vol. 22, no. 4, pp. 679–688, 2006.
- [19] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature," *Geosci. Model Dev*, vol. 7, pp. 1247–1250, 2014.

APPENDIX A

Case	HDS	HL	N <sub>HL</sub>	BS	TDS
0 (Base Case)	10	2	50%	10	80%
1	5	2	50%	10	80%
2	20	2	50%	10	80%
3	5	2	50%	5	80%
4	10	2	50%	5	80%
5	20	2	50%	5	80%
6	5	2	50%	20	80%
7	10	2	50%	20	80%
8	20	2	50%	20	80%
9	5	3	50%	20	80%
10	10	3	50%	20	80%
11	20	3	50%	20	80%
12	5	1	50%	20	80%
13	10	1	50%	20	80%
14	20	1	50%	20	80%

Case	HDS	HL	N <sub>HL</sub>	BS	TDS
15	5	2	50%	20	50%
16	10	2	50%	20	50%
17	20	2	50%	20	50%
18	5	2	100%	5	80%
19	10	2	100%	5	80%
20	20	2	100%	5	80%
21	5	2	100%	10	80%
22	10	2	100%	10	80%
23	20	2	100%	10	80%
24	5	2	100%	20	80%
25	10	2	100%	20	80%
26	20	2	100%	20	80%